09/462127
430 Rec'd PCT/PTO 29 DEC 1999

DESCRIPTION

SPEECH DECODER AND SPEECH DECODING METHOD

Technical Field

The present invention relates to a speech decoder and speech decoding method used in speech CODECs.

Background Art

Audio decoders which generate excited signals from coded speech signals input in units of frames and generate decoded speech signals from these excited signals are known. Of these types of speech decoders, in those which are adapted to low bit rate speech CODECs, the excited signals are treated with emphasis processing such as pitch emphasis processing or formant emphasis processing in order to improve the subjective sound quality of the decoded speech.

However, when frame errors occur in succession, the noise components are emphasized by these emphasis processes, thereby increasing the distortion and lowering the subjective sound quality.

Disclosure of the Invention

SUB A1

The present invention has been accomplished in view of the above considerations, and has the object of offering a speech decoder and speech decoding method capable of lightening the reduction of the subjective sound quality even when frame errors occur in succession.

In order to achieve this object, the present invention offers a speech decoder which generates excited signals from coded speech signals inputted in units of frames and generates decoded speech signals from these excited signals, characterized by comprising emphasis processing means for performing an emphasis process on said excited signals; error detecting

means for detecting frame errors in said coded speech signals; counting means for counting a number of times said frame errors occurred in succession and outputting the successive error frame number; and emphasis process prohibiting means for prohibiting said emphasis process due to said emphasis processing means when said successive error frame number exceeds a predetermined reference error frame number.

According to this speech decoder, an emphasis process is performed on the excited signals when the communication environment is good, and the successive error frame number is less than or equal to a predetermined reference error frame number. As a result, good decoded speech signals with high subjective sound quality are obtained. On the other hand, if the communication environment becomes bad and the successive error frame number exceeds the reference error frame number, the emphasis processing of the excited signals is prohibited. Therefore, distortions in the decoded speech signals which occur when emphasis processing is performed in such cases can be avoided before they occur.

Additionally, aside from prohibiting emphasis processing of excited signals when the successive error frame number has exceeded the reference error frame number, it is possible to control the amount of emphasis in the emphasis process in accordance with the successive error frame number.

Brief Description of the Drawings

Fig. 1 is a block diagram showing the structure of a speech decoder which is an embodiment of the present invention.

Fig. 2 is a block diagram showing a specific structure applying the same embodiment to a CS-ACELP type speech decoder.

Fig. 3 is a diagram for explaining a first modification example of this embodiment.

Fig. 4 is a diagram for explaining a second modification example of this embodiment.

Best Modes for Carrying Out the Invention

Next, a preferred embodiment of the present invention shall be described with reference to the drawings.

Fig. 1 is a block diagram showing the structure of a speech decoder 10 which is an embodiment of the present invention.

This speech decoder 10 comprises a decoding processing portion 11 and a emphasis process control portion 12.

SUB A2

Here, the decoding processing portion 11 is a device for decoding the received decoded speech signals (bitstream) BS and outputting the decoded speech signals SP.

This decoding processing portion 11 comprises an emphasis processing portion 15, a first switch SW1 and a second switch SW2.

The emphasis processing portion 15 performs emphasis processing with respect to the signals to be processed SPC based on the various parameters contained in the decoded speech signal, and outputs the resulting emphasized signals to be processed SEPC.

The first switch SW1 and second switch SW2 are switches for switching the signals to be processed SPC so as to be supplied to the latter-stage circuits through the emphasis processing portion 15, or so as to be supplied to the latter-stage circuits through the bypass BP.

Next, the emphasis process control portion 12 is a device for controlling whether or not to perform the emphasis processes in the decoding processing portion 11 based on frame error conditions of the coded speech signal BS.

This emphasis process control portion 12 comprises an error detecting portion 16 and a counter portion 17.

Here, the error detecting portion 16 is a device for detecting the frame errors of the coded speech signal BS and outputting error detection signals SER.

Additionally, the counter portion 17 counts the successive frame error number based on the error detection signals SER, and outputting an emphasis process control signal CE for switching the first switch SW1 and the second switch SW2 to the bypass BP side to prohibit emphasis processing when the successive frame error number exceeds a preset reference successive frame error number.

Next, the operations of the present embodiment will be described.

First, when the successive frame error number outputted from the counter portion 17 is less than or equal to a preset reference successive frame error number, the first switch SW1 and second switch SW2 are set to the emphasis process portion 15 side. Therefore, signals to be processed SPC generated from various parameters contained in the coded speech signal BS are supplied to the emphasis processing portion 15 of the decoding processing portion 11 via the first switch SW1 for emphasis processing. Then, the emphasized signals to be processed SEPC obtained by this emphasis process are outputted to the latter connected devices. As a result, a decoded speech signal SP with good subjective sound quality is obtained.

On the other hand, when the communication quality is degraded and the successive frame error number outputted from the counter portion 17 exceeds the reference successive frame error number, the first switch SW1 and second switch SW2 are set to the bypass BP side. As a result, the signals to be processed SPC generated by the parameters contained in the coded speech signal BS are outputted to latter-connected devices without being emphasis processed by the emphasis processing portion 15. Since the emphasis process is prohibited in this way when the successive frame error number is large, it is possible to reduce distortions generated by in the decoded speech signals SP.

Next, with reference to Fig. 2, a specific example of application of the present embodiment to a speech decoder in a CS-ACELP (Conjugate Structure Algebraic Code Excited Linear Prediction) type CODEC shall be explained. This type of CS-ACELP format speech coder and speech decoder are described, for example, in R. Salam et al., "Design and Description of CS-ACELP: A Toll Quality 8kb/s Speech Coder", IEEE Trans. on Speech and Audio Processing, vol. 6, no. 2, March 1998.

In Fig. 2, the speech decoder 20 comprises a parameter decoder 21. This parameter decoder 21 is a device decoding a pitch delay parameter group GP, a cobebook gain parameter group GG, a codebook index parameter group GC and an LSP (Line Spectrum Pair) index parameter group GL from the received coded speech signals (bitstream) BS.

Here, the codebook index parameter group GC includes a plurality of codebook index

parameters and a plurality of codebook code parameters.

Additionally, the speech decoder 20 comprises an adaptive code vector decoder 22, a fixed code vector decoder 23 and an adaptive preprocessing filter 25.

Here, the adaptive code vector decoder 22 is a device for outputting an adaptive code vector ACV corresponding to the pitch delay parameter group GP. More specifically, this adaptive code vector decoder 22 has a rewritable memory, and this memory contains a predetermined number of adaptive code vectors ACV which have been input in the past. The adaptive code vector decoder 22 takes the pitch delay parameter group GP as an index, reads an adaptive code vector ACV corresponding to this index from the memory, and outputs the result. Additionally, when the excited signal SEXC is reconstructed by the excited signal reconstruction portion 27 to be described later, this excited signal SEXC is written into the memory of the adaptive code vector decoder 22 as a new adaptive code vector ACV, and the oldest adaptive code vector ACV in the memory is eliminated.

The fixed code vector decoder 23 is a device for outputting an original fixed code vector FCV0 corresponding to the codebook index parameter group GC.

The adaptive code vector decoder 22 and the fixed code vector decoder 23 correspond to the codebook decoder 18 in Fig. 1.

SUB A3

The adaptive preprocessing filter 25 is a device which functions as an emphasizing process means for emphasizing the harmonic components of the decoded original fixed code vector FCV0, and outputs the result as a fixed code vector FCV.

Here, the first switch SW1 is provided in front of the adaptive preprocessing filter 25 in order to switch whether to supply the original fixed code vector FCV0 outputted from the fixed code vector decoder 23 to be supplied to the adaptive preprocessing filter 25 or to be supplied to the bypass BP. Additionally, the second switch SW2 is provided after the adaptive preprocessing filter 25 to select either the output terminal of the adaptive preprocessing filter 25 or the bypass BP for connection to the excited signal reconstruction portion 27. The first switch SW1 and second switch SW2 are switched by means of a preprocessing control signal CPR to be described later.

Furthermore, the speech decoder 20 comprises a gain decoder 24 and an LSP reconstruction portion 26.

The gain decoder 24 is a device for outputting an adaptive codebook gain ACG and a fixed codebook gain FCG based on a fixed code vector FCV (or original fixed code vector FCV0) and a codebook gain parameter group GG.

The LSP reconstruction portion 26 is a device for reconstructing the LSP coefficient CLSP based on the LSP index parameter group GL.

Further, the speech decoder 20 comprises an excited signal reconstruction portion 27, an LP synthesis filter 28, a postprocessing filter 29 and a bypass filter / upscaling portion 30.

Here, the excited signal reconstruction portion 27 is a device for reconstructing the excited signal SEXC based on adaptive code vector ACV, an adaptive codebook gain ACG, a fixed codebook gain FCG and a fixed code bector FCV (or original fixed code vector FCV0). This excited signal SEXC is written into the memory of the adaptive code vector decoder 22 as a new adaptive code vector ACV, and the oldest adaptive code vector ACV in the memory is eliminated.

The LP synthesis filter 28 is a device which performs an LP synthesis based on the excited signal SEXC and the LSP coefficient CLSP to reconstruct the speech signal SSPC.

The postprocessing filter 29 is a device for performing postprocess filtering of the speech signal SPC. This postprocessing filter 29 is constructed of three filters, a long-term postprocessing filter, a short-term postprocessing filter and a slope compensation filter. These three filters are serially connected in the order of long-term posprocessing filter to short-term postprocessing filter to slope compensation filter in the direction of input to output.

The bypass filter / upscaling portion 30 is a device for performing a bypass filtering process and an upscaling process with respect to the output signals of the postprocessing filter 29.

Additionally, the speech decoder 20 comprises an error detecting portion 31 and a counter portion 32.

Here, the error detecting portion 31 detects frame errors in the received coded speech

signals BS and outputs error detection signals SER.

Additionally, the counter portion 32 counts the successive frame error number based on the error detection signal SER, outputs a preprocessing control signal CPR for selecting the preprocessing filter 25 by means of the first switch SW1 and the second switch SW2 when the successive frame error number is less than or equal to a predetermined reference frame error number, and outputs a preprocessing control signal CPR for selecting the bypass BP by means of the first switch SW1 and the second switch SW2 when the successive frame error number has exceeded the predetermined reference frame error number.

Next, the operations of the speech decoder 20 shall be explained.

First, when the successive frame error number is less than or equal to the reference frame error number, the counter portion 32 switches the first switch SW1 and second switch SW2 to the adaptive preprocessing filter 25 by means of a preprocessing control signal CPR. As a result, the original fixed code vector FCV0 outputted from the fixed code vector decoder 23 is supplied to the adaptive preprocessing filter 25. Then, an emphasis process for emphasizing the harmonic components is performed on the original fixed code vector FCV0 in the adaptive preprocessing filter 25, and the resulting fixed code vector FCV is supplied to the gain decoder 24 and the excited signal reconstruction portion 27. Thus, a decoded speech signal SP with good subjective sound quality is obtained.

On the other hand, when the communication quality degrades and the successive frame error number outputted from the counter portion 32 exceeds the preset reference successive frame error number, the first switch SW1 and the second switch SW2 are set to the bypass BP side. As a result, the original fixed code vector FCV0 outputted from the fixed code vector decoder 23 is supplied to the gain decoder 24 and excited signal reconstruction portion 27 without undergoing an emphasis process by means of the adaptive preprocessing filter 25. Since the emphasis process is prohibited in this way when the successive frame error number is large, it is possible to reduce distortion which is generated in the decoded speech signal SP.

An embodiment of the present invention has been explained above, but various

examples of modifications to this embodiment can be considered.

Fig. 3 is a block diagram showing the structure of a speech decoder according to a first modification example. In Fig. 3, the parts which are the same as those in Fig. 1 are indicated by the same reference numerals.

In the above-described embodiment, emphasis processing is prohibited when the successive frame error number exceeds the predetermined reference successive frame error number. In contrast, in a speech decoder 30 according to a first modification example, the degree of the emphasis processing is controlled by controlling the filter gain of the preprocessing filter 25' for performing emphasis processing as shown in Fig. 3. That is, the counter portion 17' counts the successive frame error number, outputs a gain control signal SGC which makes the filter gain of the preprocessing filter 25' a normal value when this successive frame error number is less than or equal to a predetermined reference frame error number, and outputs a gain control signal SGC for making the filter gain of the preprocessing filter 25' less than usual when the successive frame error number exceeds the predetermined reference frame error number.

In this case as well, it is possible to reduce the distortions which are generated by performing emphasis processing when frame errors occur in succession, so as to enable the degradation of the subjective sound quality to be reduced.

Fig. 4 is a block diagram showing the structure of a speech decoder according to a second modification example. In Fig. 4, the parts which are the same as those in Fig. 1 are indicated by the same reference numerals.

In the speech decoder 40 of the second modification example, the deoding processing portion 41 is provided with a plurality of preprocessing filters 25'-1 to 25'-n, a first multiplexer MX1 and a second multiplexer MX2 as shown in Fig. 4.

Here, the amount of emphasis (e.g., corresponding to the filter gain) of the emphasis process performed by each of the preprocessing filters 25'-1 to 25'-n are different, the amount of emphasis in the preprocessing filter 25'-1 being the highest, and the amount of emphasis becoming lower in advancing to preprocessing filter 25'-2, preprocessing filter 25'-3 and so on.

Between the first multiplexer MX1 and the second multiplexer MX2, one route is selected from among these preprocessing filters 25'-1 to 25'-n and the bypass BP.

The counter portion 17" counts the number of successive frame errors, and supplies a selection signal SSEL for selecting the bypass BP or a preprocessing filter of an emphasis amount suited to the number of successive frame errors to the first multiplexer MX1 and the second multiplexer MX2.

In this second modification example, e.g. when the successive frame error number is "0", the preprocessing filter 25'-1 with the highest amount of emphasis is selected by the first multiplexer MX1 and second multiplexer MX2.

Then, if the communication environment worsens, preprocessing filters with lower amounts of emphasis are chosen such as preprocessing filter 25'-2 preprocessing filter 25'-3, . . . as the successive frame error number increases from "0" to "1", "2", . . .

In this way, the effects of switching of emphasis processing can be reduced because the amount of emphasis of the emphasis process can be switched in multiple steps in accordance with the successive frame error number.

In the above description, a case of a CS-ACELP type speech decoder was given as a specific example of the speech signal processing device. However, the present invention can be applied to speech signal processing devices of other formats such as speech decoders using APC (Adaptive Predictive Coding), APC-AB (APC with Adaptive Bit allocation), APC-MLQ, ATC (Adaptive Transform Coding), MPC (Multi Pulse Coding), LPC (Linear Prediction Coding), RELP (Residual Excited LPC) CELP (Code Excited LPC), LSP (Line Spectrum Pair Coding) or PARCOR as long as they are speech signal processing devices which perform emphasis processing.